Energy, Mines and Resources Canada | Énergie, Mines et Ressources Canada

CANMET

Canada Centre for Mineral and Energy Technology | Centre canadien de la technologie des minéraux et de l'énergie

HOUSEHOLD FORTRAN COMPUTER PROGRAMS FOR THE ORE PROCESSING LABORATORY: V - INTMLR, Interactive Multiple Linear Regression

J. Leung

Mineral Processing Laboratory

April 1982

MICROMEDIA

OCT 27 1982

Project MRP - 3.4.0.0.01

Mineral Processing Plant Simulation

This document is an unedited interim report prepared primarily for discussion and information reporting purposes. It does not represent a final expression of the opinion of the Canada Centre for Mineral and Energy Technology (CANMET)

Ce document est un rapport [illegible] rédigé [illegible] fin de [illegible] documentation [illegible] nullement l'opinion définitive de [illegible] Centre canadien de la technologie des minéraux et de l'énergie (CANMET)

37pp

MINERALS RESEARCH PROGRAM

MINERAL SCIENCES LABORATORIES

DIVISION REPORT MRP/MSL 82-48(IR)

CONTENTS

		Page
ABSTRACT		iii
1.0	PROGRAM IDENTIFICATION	1
2.0	ENGINEERING DOCUMENTATION	2
2.1	Narrative Description	2
2.2	Method of Solution	2
2.3	Program Capabilities	7
2.4	Data Inputs	7
2.5	Program Options	7
2.6	Printed Output	7
2.7	Flow Chart	8
2.8	Sample Run	8
3.0	SYSTEM DOCUMENTATION	8
3.1	Computer Equipment	8
3.2	Peripheral Equipment	8
3.3	Source Program	8
3.4	Variables and Subroutines	8
	Preprossessor MREGDR	8
	Main driver	9
	Subroutine INTMLR	9
	Subroutine MREG1	10
	Subroutine EXPVAR	10
	Subroutine ERRVAR	11
	Subroutine MREG4	12
	Subroutine MREG5	13
	Subroutine MREG6	13
	Subroutine MREG7	14
	IBM scientific subroutine package	15
	Library function CAS010	15
3.5	Storage Requirements	16
4.0	OPERATING DOCUMENTATION	16
4.1	Operator Instructions	16
4.2	Operating Messages	16

Table of Contents (Cont'd)

Page

4.3 Control Cards 16
4.4 Error Recovery 17
4.5 Run Times 17
REFERENCES 17

TABLES

No.
1. Key variables in INTMLR 18

FIGURES

No.
1. General flow chart for INTMLR 20
2. CCL procedure to execute the preprossessor MREGDR 22

APPENDIX

Sample runs of INTMLR 22

HOUSEHOLD FORTRAN COMPUTER PROGRAMS FOR THE ORE PROCESSING LABORATORY: V - INTMLR, INTERACTIVE MULTIPLE LINEAR REGRESSION

by

J. Leung*

ABSTRACT

INTMLR solves the problems in multi-variable linear regression. The usual (unweighted) least squares analysis is commonly used, assuming that all the observations are of equal precision. However, when some of the observations are not as reliable as the others, a weighting factor for each observation has to be taken into consideration. This interactive program allows conversational input of regression data and error variance (as a weighting factor) of each observation provided that the latter is known to the user. If there are repeat measurements of the observation made at the same values of independent variables, the "lack of fit" test can be performed to check whether the model is correct. The results of the analysis are in modular forms so that the user has the choice to either print part or all of the results.

---

* Physical Scientist, Mineral Dressing Section, Mineral Processing Laboratory, Mineral Sciences Laboratories, CANMET, Department of Energy, Mines and Resources Canada, Ottawa, K1A 0G1.

This is the fifth report of the household FORTRAN programs for the ore processing laboratory. The previous four reports that have been published separately are:

1. Laguitton, D. "METCAL, metallurgical calculations from laboratory and pilot plant investigation results"; division report MRP/MSL 79-120 (IR).

2. Laguitton, D. "PARTSZ, a program to convert particle size distribution to various units"; division report MRP/MSL 80-41 (IR).

3. Leung, J. and Laguitton, D. "INTDS, interactive driver for an optimization program with direct search method"; division report MRP/MSL 80-115 (IR).

4a). Feader, W. and Laguitton, D. "CONPLOT, a conversational program to create graphs on the CALCOMP plotter part A - system description"; division report MRP/MSL 81-10 (IR).

4b). Feader, W. and Laguitton, D. "CONPLOT, a conversational program to create graphs on the CALCOMP plotter part B - users' guide"; division report MRP/MSL 81-30 (IR).

1.0 PROGRAM IDENTIFICATION

PROGRAM TITLE: Interactive Multi-variable Linear Regression

PROGRAM CODE NAME: INTMLR

WRITER: J. Leung

ORGANIZATION: Mineral Processing Laboratory, CANMET (552 Booth Street, Ottawa, K1A 0G1)

DATE: Febuary, 1982

UPDATES: None. Version 0

SOURCE LANGUAGE: CDC FORTRAN Extended 4.8 (American National Standards Institute FORTRAN X3.9-1966)

AVAILABILITY: Complete program listing is available from the author
CANMET
Department of Energy, Mines and Resources
552 Booth Street
Ottawa, Ontario K1A 0G1

2.0 ENGINEERING DOCUMENTATION

2.1 NARRATIVE DESCRIPTION

The multi-linear analysis not only finds the regression coefficients, their standard deviations, analysis of variance table, etc. as most of the commercially available regression programs, it also tests the "lack of fit" to check whether the model is adequate to a given set of data. In simple cases, this can be detected merely by examining a plot of the data. When the model is more complicated and/or involves more variables, this may not be possible. In order to see whether there is lack of fit, an estimate of the variance of the observation has to be obtained either from previous experience of the variation in the situation being studied or from the repeat measurements of Y made at the same values of X. The latter estimate is much more reliable than any other source. The usual procedure is to test the F ratio (see METHOD OF SOLUTION).

It may happen that some of the observations used in regression analysis are not as reliable as the others. In other words, the variances of the observations are not the same. In practical problems it is often difficult to obtain the complete variance matrix. In this work, we assume that the variances of the observations are not correlated, i.e., the off-diagonal elements of the variance matrix are zero. Even then, the diagonal terms are sometimes not available at first. In this case, it is necessary to make the assumption that it is a "unity" matrix and one is left to examine the residuals from the regression analysis.

To execute INTMLR routine, a preprossessor called MREGDR has to be run first. MREGDR reads in the numbers that are required to set the dimensions of the variables in the program; writes these numbers in a driver which serves as a main program for INTMLR and passes the dimensions through arguments. The operator-controlled predimensioning reduces the unnecessary computer storage memory.

2.2 METHOD OF SOLUTION

The least squares method is used to find the regression coefficients of a multi-variable linear regression in the form:

$$Y = \beta_o + \beta_1 (X_1 - \bar{X}_1) + \ldots + \beta_p (X_p - \bar{X}_p) + \epsilon = \hat{Y} + \epsilon$$

where $\bar{X}_1, \ldots, \bar{X}_p$ are the mean values of the independent variables $X_1, \ldots, X_p$ respectively, $\beta_o, \beta_1, \ldots, \beta_p$ are the regression coefficents, Y is the corresponding observation and ϵ is the increment by which any individual Y may fall off the regression point $\hat{Y}$. The derivation of the parameters by minimizing the sum of squares of the deviations can be found in many statistical books and reports (1,2,3) and it is not the purpose of this manual. However, the readers are expected to understand some basic statistical terminology before going further into this section.

One main concern is how to justify that the model adequately fits the data by the results of the regression analysis. The most common criteria are:

1. multi-correlation coefficient
2. lack of fit test
3. fraction of the corrected sum of squares due to the regression R^2.

Normally, lack of fit test is the most powerful of all three concepts. When the values calculated for 1 and 3 are > 0.95, the criterion of "good fit" is met. The former value simply implies that there is a good association between the values of observed and predicted Y and does not by itself imply any sort of relationship between them. The latter value measures the proportion of total variation about the mean value of Y contributed by the regression. Therefore the results do not give us enough information on the model itself.

Certain basic assumptions are made in the above regression equation. The error term ϵ which could be a sum of errors from several sources tends to be normally distributed due to the Central Limit theorem (4). Thus we assume that ϵ_i (for ith observation) is a random variable with mean zero and variance σ^2, that is,

$$\sum_{i=1}^{n} \epsilon_i = 0$$

$$V(\epsilon_i) = \sigma^2$$

and is uncorrelated with each other. If the model is correct, then the residual mean square S_r^2 is an estimate of σ^2. (Its square root is called the standard error of estimate in the program.) If the postulated model is not the true model, then $\sigma^2 < S_r^2$ and the postulated model is said to be suffering from lack of fit. In some cases, an estimate of σ^2 can be obtained from previous experience of the same situation and we shall call it σ_y^2. The function (3)

$$M = \frac{\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2}{\sigma_y^2} = \frac{(n-p-1)S_r^2}{\sigma_y^2}$$

where $\hat{Y}_i$ is the predicted Y_i value and p the number of independent variables, follows a χ^2-distribution with n-p-1 degrees of freedom. If, for a given level of significance α, the quantity M exceeds $\chi^2_{1-\alpha}$, we say that the χ^2-test fails and the model is inadequate in its present form. The error ϵ_i might contain the systematic error (or bias error) besides the random error.

The more reliable estimate of σ^2 is the experimental variance S_e^2, measured from the repeat measurements of Y made at the same values of X. The mean squares due to the lack of fit is

$$S_m^2 = \frac{N_r S_r^2 - N_e S_e^2}{N_r - N_e}$$

where N_r and N_e are the numbers of degrees of freedom of S_r^2 and S_e^2 respectively. The ratio $F = S_m^2/S_e^2$ is compared with the $100(1-\alpha)\%$ point of an F distribution with (N_r-N_e) and N_e degrees of freedom where (N_r-N_e) is the number of degrees of freedom of S_m^2. If the ratio is significant, it indicates that the model is inadequate and attempts to discover the cause of inadequacy can be started by plotting the residuals (1). If the ratio is not significant, the model can be retained and S_r^2 is used as an estimate of σ^2.

At least one of the tests mentioned here should be performed before using any of the results calculated by the program. The covariance matrix of

the parameters and the confidence intervals of the parameters and the predicted Y values are estimated with S_r^2 in INTMLR routine under the assumption that the model is retained. Therefore, it is the responsibility of the user to do the test with the information provided by the program and design a new model if the existing model is not adequate.

There are some ways to check the significance of the parameters in a model with the following calculations:

1. computed t-values
2. F-ratio
3. confidence intervals.

The computed t-value of each parameter calculated in the program should exceed $t\ (n-p-1,\ 1 - \frac{\alpha}{2})$ from a t-table if that parameter is to be retained in the model (1). The F-ratio obtained in the analysis of variance table is used to test the null hyporthesis that the true value of overall regression is zero (more specifically, test $\beta_1 = \beta_2 = \ldots = \beta_p = 0$ against not all $\beta_i = 0$). It should be larger than the $100(1-\alpha)\%$ point of F-distribution with p and (n-p-1) degrees of freedom in order to reject the hyporthesis. The fact that the F-ratio is "significant" does not mean that the equation is useful for predictive purposes unless the range of values predicted by the fitted equation is considerably greater than the size of the random error (1). Test for individual regression coefficients which can also be made by a similar argument is not included in this work. The other way of checking whether a given parameter could be omitted from the model is to determine whether the zero point lies within the confidence interval of the given parameter or not. If it does not, the parameter is said to be significantly different from zero and should be retained.

In many cases the observations Y_i are not all of equal precision and if σ_{yi}^2 is a measure of the expected deviation from the true value for Y_i, the obvious generalization of the least squares principle is to minimize

$$\sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2/\sigma_{yi}^2$$

In other words, the observations are of different weights and the variance matrix $V(\epsilon)$ is diagonal with unequal diagonal elements. It may also

happen, in some problems, that the off-diagonal elements of $V(\varepsilon)$ are not zero, that is, the observations are correlated. This is beyond our considerations in this work. The mean value of each independent variable is defined as

$$\bar{X}_i = \frac{\sum_{j=1}^{n} X_{ji}/\sigma_{yj}^2}{\sum_{j=1}^{n} 1/\sigma_{yj}^2}$$

and the minimum function becomes

$$M = \sum_{i=1}^{n} \frac{(Y_i - \hat{Y}_i)^2}{\sigma_{yi}^2}$$

Matrix V (ε) is then used to find the variance matrix of the parameters and each diagonal element is treated as an estimator of ε to calculate the confidence interval on measured Y for different values of X.

Sometimes, the variance of Y_i is proportional to Y_i, for instance

$$\sigma_y^2 = kY^2$$

where k is the unknown proportionality factor. If the model is correct, k can be estimated by the following

$$k = \frac{1}{n-p-1} \sum_{i=1}^{n} \frac{(Y_i - \hat{Y}_i)^2}{Y_i^2}$$

and so is the error variance σ_y^2 of each Y. In this work, σ_y^2 can be either a numerical value or any function of the variables, for example,

$$0.5*X_1 + X_2^2 \text{ or } kX_1^2$$

If an unweighted least squares analysis is performed on observations of different precisions, the estimates obtained are still unbiased but do not have minimum variance, since the minimum variance estimates are only obtained from the correct weighted least squares analysis.

2.3 PROGRAM CAPABILITIES

The maximum numbers of observations NN1 and independent variables NN2 are entered prior to executing MREGDR. This enables the user to enter appropriate dimensions for each problem and thus reduces memory requirements.

2.4 DATA INPUTS

As mentioned above, the dimensions for NN1 and NN2 have to be entered at the beginning of the run. Once the driver has been written, data are then input interactively in free format separated by a comma or a blank (refer to sample runs).

2.5 PROGRAM OPTIONS

The program will perform "lack of fit" test and weigh each observation if the user answers yes respectively to the following two questions:

1) Any repeat Y for same X values?
2) Do you know the error variance of Y?

The former question will then lead to other questions about the positions of the repeat measurements (see sample runs). The error variance can either be a numerical value or a function of any valid FORTRAN expression. Usually the later is proportional to a known variable. The program will estimate the proportionality factor K if it is unknown (see also sample runs).

2.6 PRINTED OUTPUT

Basic results of the regression analysis are printed. The test on "lack of fit", the value of the unknown proportionality factor K and the error variance derived for each Y are also printed if the user has input the required data. The operator has the choice on the following outputs: table of residuals, table of variance and 95% confidence intervals for each parameter and estimated Y value.

2.7 FLOW CHART

The flow chart of INTMLR is shown in Fig. 1.

2.8 SAMPLE RUN

Three sample runs are shown in the Appendix. Each shows different options of the program.

3.0 SYSTEM DOCUMENTATION

3.1 COMPUTER EQUIPMENT

INTMLR runs on a CDC 730 Cyber computer with a maximum of 77K of core memory available for time sharing jobs.

3.2 PERIPHERAL EQUIPMENT

A time-sharing terminal is necessary.

3.3 SOURCE PROGRAM

The source listing is availabe from the author, CANMET, 552 Booth St., Ottawa, Canada.

3.4 VARIABLES AND SUBROUTINES

The key variables are defined in Table 1.

Preprossessor - MREGDR

Purpose

To read in the dimensions NN1 and NN2 and write a driver on tape 13 to execute INTMLR.

Sequence of steps

1- Set input and output file number
2- Rewind output file
3- Read NN1, NN2
4- Set NN3 = NN2 + 1
5- Write the driver in FORTRAN
6- Rewind output file

Main Driver

The driver is written and stored in tape 13 after the execution of MREGDR. To proceed the regression analysis, one has to compile and execute tape 13.

Purpose

Serve as main program to call INTMLR and pass the values of NN1, NN2 and NN3 as well as all variables storage locations to INTMLR.

Internal calls

Call INTMLR

Subroutine INTMLR

Purpose

Act as if it is a main program to call other subroutines to perform regression analysis.

Sequence of steps

1- Call MREG1 for data input
2- Check whether there are repeat Y for same X values.
   If yes call EXPVAR
3- Check whether any input error variance for Y.
   If yes call ERRVAR
4- Call MREG4 to calculate parameters, covariance matrix and Y (estimate)
5- Call MREG5 for correlation coefficients
6- Call MREG6 to print default output or table of residuals (optional)
7- Call MREG7 to perform analysis of variance and find 95% confidence intervals on parameters and Y (estimate)
8- Return to step 1 if so desired
9- Return

Internal calls

Call MREG1,EXPVAR,ERRVAR,MREG4,MREG5,MREG6 and MREG7.

DIMENSION statements

Dimensions of all variables are represented by NN1, NN2 and NN3. The actual values are entered prior executing MREGDR and passed to INTMLR through arguments transmission.

Calling arguments

Call INTMLR (NN1,NN2,NN3,X,Y,VAY,YE,X1.DIAB,TCOM,XBAR,B,YEBAR,XTV1X, SU,NREP,NR,YBAR1,LR,MR,XTX,W,XTV1,SX2,SXY2,YYBA,A,AT,AX)

Subroutine MREG1

Purpose

To input values of NOBV, NIDV, X and Y

Calling arguments

Call MREG1 (NI,NO,NN1,NN2,NOBV,NIDV,X,Y)

Description of arguments

NI: Input file number

NO: Output file number

NN1: Dimension set for NOBV

NN2: Dimension set for NIDV

NOBV: Number of observations

NIDV: Number of independent variables

X: X array with (NOBV) rows and (NIDV) columns

Y: Y vector with (NOBV) rows

Subroutine EXPVAR

Purpose

To calculate the experimental variance

Calling arguments

Call EXPVAR (NI,NO,NN1,Y,SE2,NLL,SU,NREP,NR,YBAR1)

Description of variables

SE2: Experimental variance of Y

NLL: Degrees of freedom of SE2

NRS: Number of different values of X for which there is repeat

NREP: Number of different values of Y for each repeat of X

NR: Positions of different Y for same X values (start with the first repeat of X)

YBAR1: Mean value of Y in each repeat of X

SU: Sum of squares of (Y-YBAR1) over (NRS) repeats

Subroutine ERRVAR

Purpose

To read error variance of each observation. It can either be numerical value or a function of known variables. (The latter is required to be entered by the user once only.)

Calling arguments

Call ERRVAR (NI,NO,NN1,NN2,NOBV,X,Y,VAY,LOGVFN)

Description of variable

VAY: Output vector of error variance

(Refer to table 1 for variables not defined here)

Sequence of steps

1- Check whether error variance is a function of variables

2- If not, enter error variance of each Y

3- Return

4- If yes, read error variance function in terms of V1, V2, ..., V10 once only

5- Define X,Y-variables with ten V-variables (X1 to X9 represented by V1 to V9 and Y by V10. for other X-variables, it is the responsibility of the user to modify this part of the subroutine)

6- In case of VAY=K*(FUNCTION) where K is the proportionality factor, check whether K is known. (If not, K will be estimated in subroutine MREG4)

7- Return

Function required

CAS010

Subroutine MREG4

Purpose

To perform multi-linear regression analysis

Calling arguments

Call MREG4 (NN1,NN2,NN3,X,Y,NOBV,NIDV,VAY,LOGVFN,X1,YE,CMRES,B,XBAR,
XTV1X,SYEY2,LR,MR,XTX,W,XTV1,CNST,LOGVAY)

Description of variables

X1: (X-XBAR) matrix with dimensions NN3 x NN3

NN3: NN2 + 1

LR,MR: Work vector of length NN3

XTX: Work matrix with dimensions NN3 x NN1

XTV1: Work matrix with dimensions NN3 x NN1

CNST: Proportionality factor in error variance function if any

W: Weighting factor for each Y

(Refer to table 1 for variables not defined here)

Sequence of steps

1- Calculate mean or weighted mean value XBAR of each X
2- Define (X-XBAR) matrix and convert it to single dimension required in IBM subroutines
3- Call IBM scientific subroutines for matrix manipulations
4- Find YE, CMRES, XTV1X and B
5- Estimate CNST if unknown and then VAY
6- RETURN

Internal calls

Call IBM subroutines: ARRAY, TPRD, GMPRD, MINV, SMPY

Subroutine MREG5

Purpose

To find standard deviation of X and Y, correlation of X versus Y and multi-correlation coefficient

Calling arguments

Call MREG5 (NN1,NN2,NOBV,NIDV,X,Y,XBAR,PYYC,YEBAR,YBAR,SY,SX2,SXY2, YYBA,YE)

Description of parameters

YEBAR: (YE-YBAR) vector of length NN1

YYBAR: (Y-YBAR) vector of length NN1

(Refer to table 1 for variables not described here)

Internal calls

Call IBM subroutines: SSUB, MATA

Subroutine MREG6

Purpose

Print default output or table of residuals (optional)

Calling arguments

Call MREG6 (NI,NO,NIDV,NOBV,Y,YE,XBAR,SX,SXY,B,TCOM,LOGREP,YBAR,SY, PYYC,CMRES,SE2,NLL,XX,CNST,VAY,LOGVFN,NN1,NN2,NN3,TDIS,DIAB)

Description of parameters

NI: Input file number

NO: Output file number

XX: Covariance matrix of regression coefficients

CNST: Proportional factor in error variance function if any

DIAB: Diagonal values of XX

(Refer to table 1 for variables not described here)

Internal calls

Call IBM subroutines: DCPY, ARRAY

Sequence of steps

1- Find T-distribution function approximation
2- Convert XX from single to double dimension
3- If ABS(XX) $\leq 9*10^{-9}$, set XX=0.
4- Perform "lack of fit" test, if SE2 is available
5- Output basic results
6- Print table of residuals if so desired by the user
7- Return

Subroutine MREG7

Purpose

Find 95% confidence intervals for estimated Y and measured Y as well as the regression coefficients. The analysis of variance is also performed. The results can be printed out if wanted.

Estimated $Y = \beta_o + \beta_1(X_1-\bar{X}_1) + \ldots + \beta_p(X_p-\bar{X}_p)$

Measured Y = Estimated Y + ε

Calling arguments

Call MREG7 (NI,NO,NN1,NN3,NOBV,NIDV,YEBAR,CMRES,X1,XTV1X,TDIS,YE, DIAB,B,A,AT,AX,SYEY2,VAY,LOGVAY)

Description of parameters

A - Working vector of length NN3
AT - Transpose of A
AX - Working vector of length NN3

(Refer to subroutine MREG4, MREG5 and table 1 for variables not described here)

Sequence of steps

1- Perform analysis of variance
2- Check whether the print-out is wanted
3- Find 95% confidence intervals

4- Check whether the print-out is wanted

5- Return

Internal calls

Call IBM subroutines: GMTRA, GMPRD

IBM Scientific Subroutine Package

Purpose

To perform matrix manipulations

For details on subroutines GMTRA, GMPRD, ARRAY, TPRD, MINV, SMPY, DCPY, MATA, SSUB, refer to (5)

Library Function CAS010

Purpose

To provide a facility for processing a string containing an arithmetic Fortran expression as data.

Calling sequence/argument list

Z = CAS010 (ISTRING,NO)

To pass an input arithmetic Fortran string through array ISTRING to CAS010

(For details, refer to users guide for Applications Program Library, Computer Science Center, Department of Energy, Mines and Resources)

ISTRING: 240 work integer array which holds the input arithmetic FORTRAN string

NO: Output device number for error messages

COMMON block

COMMON/VIV10/V1,V2,V3,V4,V5,V6,V7,V8,V9,V10

CAS010 recognizes only V1, V2,...,V10 as legal names of independent variables and can handle a maximum of up to 10 only.

3.5 STORAGE REQUIREMENTS

Name	Octal Words
MREGDR	1000
DRIVER	2677
INTMLR	553
MREG1	141
EXPVAR	217
ERRVAR	617
MREG4	366
MREG5	213
MREG6	577
MREG7	543

4.0 OPERATING DOCUMENTATION

4.1 OPERATOR INSTRUCTIONS

INTMLR has been developed and tested on the CYBER730 computer of the Computer Science Center at the Dept. of Energy, Mines and Resources, under the INTERCOM operating system.

4.2 OPERATING MESSAGES

Normal system messages are produced by the INTERCOM system. No special diagnostics are issued by the program.

4.3 CONTROL CARDS

The program was run interactively. A Cyber Control Language (CCL) procedure controls automatically all ATTACH, LOAD and EXECUTE operations of appropriate disk files. One single command will execute the program:

```
BEGIN, MREGDR, MPROC/CCL, P1, P2.
```

where P1 and P2 are the dimensions the user set for the numbers of observations and independent variables respectively. The procedure is listed in Figure 2. This procedure should be replaced by its equivalent on another computer or the file manipulation should be manual.

4.4 ERROR RECOVERY

Program must be restarted on error.

4.5 RUN TIME

Run time varies with the problem size and the chosen output options. The sample runs of this program required a total of 1.2 CP seconds.

REFERENCES

1. Draper, N. and Smith, H. "Applied regression analysis", John Wiley & Sons, 1966.

2. Finch, J.A. and Laplante, A.R. "Methodology for ore and coal process modelling", SPOC project, CANMET, 552 Booth Street, Ottawa.

3. Brandt, S. "Statistical and computational methods in data analysis", North-Holland publishing Co., 1976.

4. Mood, A.M. "Introduction to the theory of statistics", McGraw-Hill Book Co., Inc., 1950.

5. IBM application program, system/360 scientific subroutine package version III, programmer's manual, program no. 360A-CM-03X, Aug. 1970.

Table 1 - Key variable description of INTMLR

Name	Dimension	Description
NOBV	1	Number of observations
NIDV	1	Number of independent variables
X	NN1 x NN2	Independent variables
Y	NN1	Dependent variable
YE	NN1	Estimated Y
VAY	NN1	Input error variance of each Y
B	NN3	Regression coefficients
XTV1X	NN3 x NN3	Covariance matrix of B
NRS	1	Number of different values of X for which there is repeat Y
NREP	NN1	Number of different values of Y for each repeat of X
NR	NN1	Positions of different values of Y for each repeat of X (start with the first repeat)
YBAR1	NN1	Mean value of Y in each repeat
SU	NN1	Sum of squares of (Y-YBAR1) over NRS repeats
VU1,VU2	1	Upper limits (95% confidence) for estimated and measured Y (see subroutine MREG7)
VL1,VL2	1	Lower limits (95% confidence) for estimated and measured Y
SE2	1	Experimental variance calculated from repeat measurements
NLL	1	Degrees of freedom of SE2
XBAR	NN2	Mean or weighted mean value of X
YBAR	1	Mean value of Y
CMRES	1	Residual mean squares
SYE2	1	Variance of estimated Y
SYSS	1	Variance of measured Y
TDIS	1	T-distribution function (95%)
SX2	NN2	Standard deviation of X

Table 1 (Cont'd)

Name	Dimension	Description
SXY2	NN2	Correlation of X versus Y
SY	1	Standard deviation of Y
PYYC	1	Multi-correlation coefficient
TCOM	NN3	Computed T-value
SREG	1	Sum of squares due to regression
SYEY2	1	Sum of squares about regression
STTL	1	Sum of squares about the mean
VUB,VLB	1	Upper and lower limits (95% confidence) for each regression coefficient
RD2	1	SREG/STTL
SM2	1	Mean squares due to lack of fit
W	NN1	Weighting factor for each Y
DIAB	NN3	Diagonal elements of XTV1X

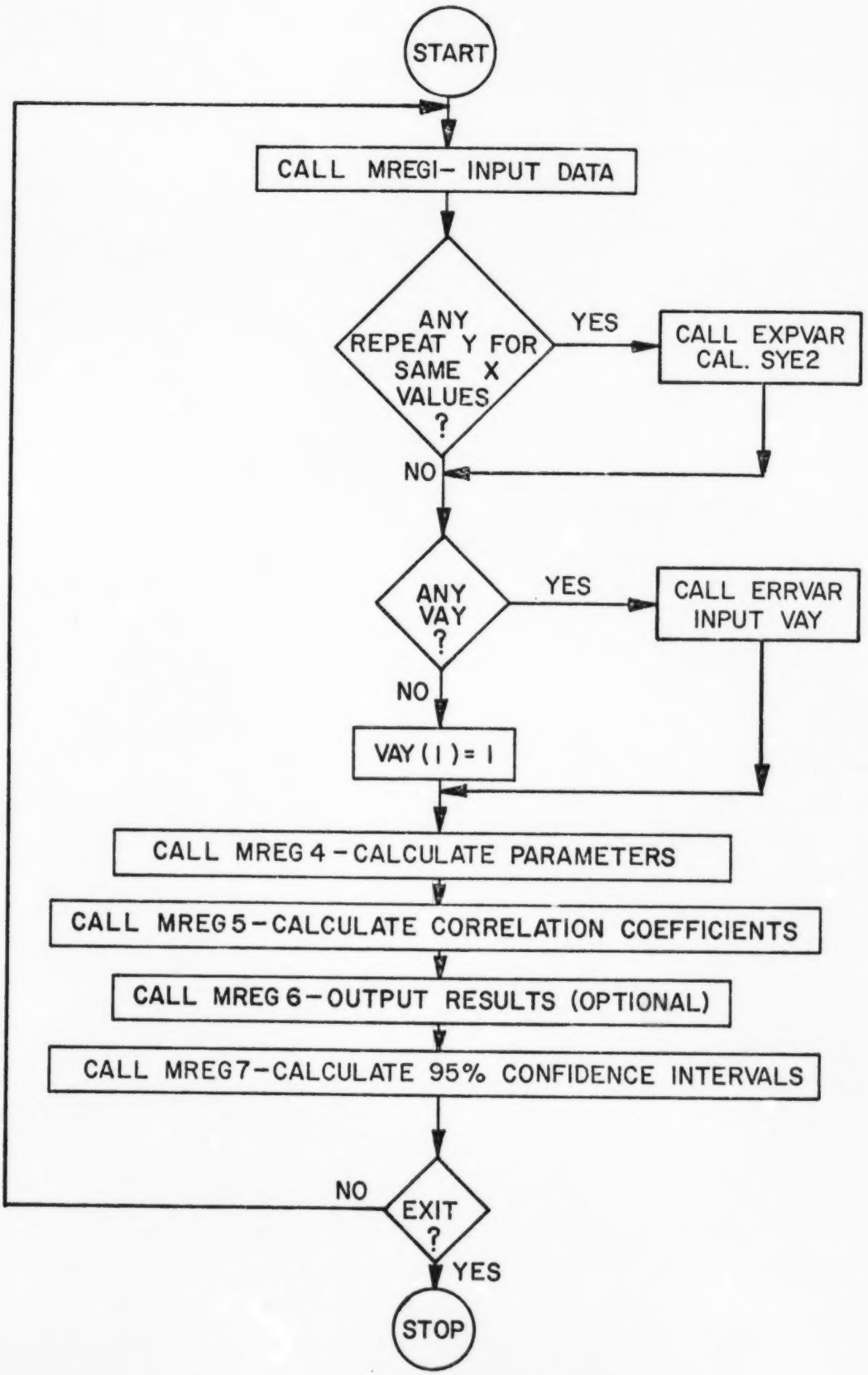

FIG. 1 FLOW CHART FOR INTMLR

```
.PROC, MREGDR, V1=, V2=.
ATTACH, MREGDR, ID=BIN, MR=1.
MREGDR.
FTN, I=TAPE13, B=LGO, L=0.
ATTACH, INTMLR, ID=BIN.
ATTACH, SSPIBM, ID=LIB, MR=1.
ATTACH, SIMLIB, ID=LIB, MR=1.
LIBRARY, SSPIBM, SIMLIB.
CONNECT, INPUT, OUTPUT.
LOAD (INTMLR).
LGO.
RETURN, MREGDR, TAPE13, INTMLR, SSPIBM, SIMLIB, LGO.
LIBRARY.
EXIT(U)
RETURN, MREGDR, TAPE13, INTMLR, SSPIBM, SIMLIB, LGO.
LIBRARY.
REVERT.
.DATA, INPUT.
V1, V2
```

Fig. 2 - CCL procedure to execute the preprossess MREGDR

APPENDIX

OUTPUT FROM SAMPLE RUNS

(These examples are taken from, "Methodology for Ore and Coal Process Modelling" by J.A. Finch and A.R. Laplant, SPOC Manual, Chapter 4)

```
                    ************** INTMLR ***************

CURVE FITTING FOR MULTI VARIABLE LINEAR REGRESSION
          Y= B0 + B1*(X1-X1BAR) + B2*(X2-X2BAR) + ...
          WHERE X1BAR, X2BAR ARE THE MEAN VALUES OF
          X1, X2 RESPECTIVELY
DATA ENTRY IS FREE FIELD FORMAT


ENTER THE NUMBER OF OBSERVATIONS, NOBV=
17

ENTER THE NUMBER OF INDEPENDENT VARIABLES, NIBV=
3

ENTER X  1 VALUES
608,554,626,597,623,559,611,628,623,579,613,562,[illegible]

ENTER X  2 VALUES
360,354,330,336,333,330,324,365,352,317,319,314,[illegible]

ENTER X  3 VALUES
77,77,67,67,69,77,78,74,72,76,68,69,69,66,71,67,72

ENTER X  4 VALUES
175,181,132,141,135,161,147,160,123,143,152,120,[illegible]


ENTER Y VALUES
2.0565,2.70,2.2257,2.7213,3.0253,2.5649,2.0549,2.7285,[illegible]
2.8024,2.8214,2.451,2.4681,2.5257,2.2618,3.054,2.954
```

DO YOU WANT REPEAT Y FOR SAME X VALUES ?
no

DO YOU KNOW THE ERROR VARIANCE FOR EACH Y ?
no

VARIABLE NO.	MEAN	STANDARD DEVIATION	CORRELATION X VS Y	REGRESSION COEFFICIENT	COMPU. T VALUE
2	600.47059	31.68027	1.75815	.05548	6.31
3	332.41176	20.02517	.41423	.00540	6.81
4	71.29412	5.89242	.75572	.04144	9.83
5	147.11765	18.42852	1.18058	-.01863	10.60
DEP.	2.77801	.24340			

T (12), 95 %: 2.18 (TWO TAILS)

INTERCEPT	2.77801
MULTI-CORRELATION COEFFICIENT	.98467
STD. ERROR OF ESTIMATE	.07403

COVARIANCE MATRIX OF REGRESSION COEFFICIENTS :

.322oE-03	0.	0.	0.
0.	.3646E-06	.9371E-07	-.4480E-06
0.	.9371E-07	.1922E-05	.1441E-05
0.	-.4480E-06	.1441E-05	.1651E-04
0.	.4551E-07	-.1747E-05	-.4184E-05

0.	.4551E-07
0.	-.1747E-05
0.	-.4184E-05
0.	.3139E-04

DO YOU WANT THE TABLE OF RESIDUALS ?yes

TABLE OF RESIDUALS

CASE NO.	Y VALUE	Y ESTIMATE	RESIDUAL
1	2.85650	2.74939	.10711
2	2.76000	2.61839	.14161
3	2.99570	2.98295	.01275
4	2.70130	2.74398	-.04268
5	3.02530	3.00494	.02036
6	2.56490	2.63330	-.06840
7	2.95490	2.99641	-.04151
8	2.92850	3.01832	-.08982
9	3.07270	3.01302	.05968
10	2.80940	2.85083	-.04143
11	2.82140	2.82798	-.00658
12	2.45100	2.41102	.03998
13	2.46810	2.42005	.04805
14	2.52570	2.64309	-.11739
15	2.26180	2.31031	-.04851
16	3.05400	2.99908	.05492
17	2.95490	2.96302	-.00812

DO YOU WANT THE ANALYSIS OF VARIANCE FOR THE MODEL ?yes

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F-RATIO
REG	.8821	4	.2205	40.3114
RES	.0658	12	.0055	
TTL	.9479	16	.0592	

R**2 (REG/TTL)= .93058

DO YOU WANT THE 95% CONFIDENCE INTERVALS FOR Y ESTIMATE AND THE PARAMETERS ?yes

95% CONFIDENCE INTERVALS FOR Y AND Y*

Y = B0 + B1*(X1-X1BAR) + B2*(X2-X2BAR) + ...

Y*= B0 + B1*(X1 X1BAR) + B2*(X2 X2BAR) +... + E

(E IS THE ERROR ON Y)

Y ESTIMATE	Y LOWER LIMIT	Y UPPER LIMIT	Y* LOWER LIMIT	Y* UPPER LIMIT
2.74939	2.53408	2.96470	2.48034	3.01844
2.65839	2.54165	2.77513	2.45924	2.85753
2.98295	2.90545	3.06045	2.80397	3.16193
2.74398	2.71042	2.77754	2.57919	2.90877
3.00494	2.94799	3.06190	2.83385	3.17604
2.63330	2.53221	2.73438	2.44291	2.82369
2.99641	2.93864	3.05418	2.82504	3.16778
3.01832	2.83223	3.18440	2.78677	3.24987
3.01302	2.84970	3.17634	2.78345	3.24259
2.85083	2.77521	2.92646	2.67265	3.02901
2.82798	2.68943	2.96653	2.61532	3.04063
2.41102	2.24252	2.57951	2.17774	2.64430
2.42005	2.33877	2.50133	2.23940	2.60070
2.64309	2.55266	2.73353	2.45814	2.82805
2.31031	2.19038	2.43024	2.10928	2.51134
2.99708	2.81209	3.18208	2.75211	3.24206
2.96302	2.90526	3.02079	2.79166	3.13439

95% CONFIDENCE INTERVALS FOR PARAMETERS

	LOWER LIMIT	UPPER LIMIT
B 0	2.73888	2.81714
B 1	.00096	.00359
B 2	.00643	.01247
B 3	.03167	.04937
B 4	-.02249	-.01477

```
DO YOU WANT TO EXIT PROGRAM #INTMLR$ ?
NO

ENTER THE NUMBER OF OBSERVATIONS, NOBV=
23

ENTER THE NUMBER OF INDEPENDENT VARIABLES, NIDV=
2

ENTER X  1 VALUES
1.0035,.7279,.8305,1.1157,1.1202,.9664,.4560,.9560,.7151,1.7111
.7110,.7106,.7074,.7074,.0275,.0270,.0270,.0270,.0110,.0107,.1500,.15

ENTER X  2 VALUES
1.0009,.5006,.7008,1.344,1.389,.9907,.1791,.6069,.0233,.0807,.1532
.4692,.4751,.5036,-.00016,.00064,.00092,.00092,.00015,.0017,.0108,.07072

ENTER Y VALUES
12.5,9.6,10.3,14.15.3,12.5,6.26,12.45,7.19,23.../2.17,9.18,7.55,7.53
.37,.37,.1,.4,.56,.56,2.16,2.16

ANY REPEAT Y FOR SAME X VALUES ?
NO

DO YOU KNOW THE ERROR VARIANCE OF Y ?
YES

IS ERROR VARIANCE A FUNCTION OF VARIABLES (E.G. A*V1+B*X1**C)
YES

ENTER VARIANCE IN TERMS OF V1 TO V10
WHERE X1 TO X9 REPRESENTED BY V1 TO V9 AND Y BY V10 (ENTER ONCE ONLY)
V10*V10

IN CASE OF ERROR VARIANCE EQUALS TO K*(ANY FUNCTION),
DID YOU ENTER THE VALUE OF PROPORTIONALITY FACTOR K ?
(ENTER YES FOR OTHER CASES WHERE A,B,... HAVE ALREADY BEEN ENTERED)
NO
```

VARIABLE NO.	MEAN	STANDARD DEVIATION	CORRELATION X VS Y	REGRESSION COEFFICIENT	COMPUTED T VALUE
2	.03542	.73594	1.66974	14.00169	93.53775
3	.00368	1.01484	2.03095	-.96953	-6.60958
DEP.	7.64455	5.93245			

T (19), 95 %: 2.09 (TWO TAILS)

INTERCEPT	.47622
MULTI-CORRELATION COEFFICIENT	.99644
STD. ERROR OF ESTIMATE	.26041

ESTIMATED PROPORTIONALITY FACTOR FOR ERROR VARIANCE .00031150

CASE NO.	ERROR VARIANCE (ESTIMATE)
1	.06429687
2	.03792384
3	.04799735
4	.08065399
5	.09507295
6	.06429687
7	.01612570
8	.06378352
9	.03705973
10	.17462412
11	.03690369
12	.03698167
13	.03729441
14	.03582062
15	.00005633
16	.00005633
17	.00006584
18	.00006584
19	.00012905
20	.00012905
21	.00191989
22	.00191989

COVARIANCE MATRIX OF REGRESSION COEFFICIENTS :

.1209E-04	0.	0.
0.	.2241E-01	-.1937E-01
0.	-.1937E-01	.2152E-01

DO YOU WANT THE TABLE OF RESIDUALS ?YES

TABLE OF RESIDUALS

CASE NO.	Y VALUE	Y ESTIMATE	RESIDUAL
1	12.50000	13.06416	-.56416
2	9.60000	9.68260	-.08260
3	10.80000	10.93283	-.13283
4	14.00000	14.28312	-.28312
5	15.20000	14.32169	.87831
6	12.50000	12.56428	-.06428
7	6.26000	6.19780	.06220
8	12.45000	12.50384	-.05384
9	9.49000	9.49429	-.00429
10	20.60000	20.36712	.23288
11	9.47000	9.46628	.00372
12	9.48000	9.59058	-.11058
13	9.52000	9.42804	.09196
14	9.33000	9.40041	-.07041
15	.37000	.36847	.00153
16	.37000	.37389	-.00389
17	.40000	.40022	-.00022
18	.40000	.40022	-.00022
19	.56000	.56676	-.00676
20	.56000	.55167	.00833
21	2.16000	2.11613	.04387
22	2.16000	2.11997	.04003

DO YOU WANT THE ANALYSIS OF VARIANCE FOR THE MODEL ?NO

DO YOU WANT THE 95% CONFIDENCE INTERVALS FOR Y ESTIMATE
AND THE PARAMETERS ?NO

DO YOU WANT TO EXIT PROGRAM *INTMLR* ?
NO

```
ENTER THE NUMBER OF OBSERVATIONS, NOBV=
5

ENTER THE NUMBER OF INDEPENDENT VARIABLES, NIDV=
1

ENTER X  1 VALUES
2,6,6,6,10

ENTER Y VALUES
4.01,5.03,4.81,5.04,8.13

ANY REPEAT Y FOR SAME X VALUES ?
YES

ENTER THE NUMBER OF DIFFERENT VALUES OF X FOR WHICH THERE IS REPEAT
1

ENTER THE NUMBER OF DIFFERENT VALUES OF Y FOR EACH REPEAT OF X
3

ENTER THE POSITIONS OF REPEAT MEASUREMENTS
(START WITH THE FIRST REPEAT OF X)
2,3,4

DO YOU KNOW THE ERROR VARIANCE OF Y ?
NO
```

VARIABLE NO.	MEAN	STANDARD DEVIATION	CORRELATION X VS Y	REGRESSION COEFFICIENT	COMPUTED T VALUE
2	6.00000	2.82843	2.74226	.94000	25.09270
DEP.	4.84400	2.66503			

T (3), 95 %: 3.18 (TWO TAILS)

INTERCEPT	4.84400
MULTI-CORRELATION COEFFICIENT	.99763
STD. ERROR OF ESTIMATE	.21191

LACK OF FIT TEST

MEAN SQUARES		DEGREES OF FREEDOM	RATIO
LACK OF FIT	.10092	1	5.97160
EXP. ERROR	.01690	2	

COVARIANCE MATRIX OF REGRESSION COEFFICIENTS :

```
  .8981E-02   0.
  0.           .1403E-02
```

DO YOU WANT THE TABLE OF RESIDUALS ?YES

TABLE OF RESIDUALS

CASE NO.	Y VALUE	Y ESTIMATE	RESIDUAL
1	.91000	1.08400	-.17400
2	5.03000	4.84400	.18600
3	4.81000	4.84400	-.03400
4	5.04000	4.84400	.19600
5	8.43000	8.60400	-.17400

DO YOU WANT THE ANALYSIS OF VARIANCE FOR THE MODEL ?YES

ANALYSIS OF VARIANCE TABLE

SOURCE	SUM OF SQUARES	DEGREES OF FREEDOM	MEAN SQUARES	F-RATIO
REG	28.2752	1	28.2752	629.6437
RES	.1347	3	.0449	
TTL	28.4099	4	7.1025	

R**2 (REG/TTL)- .99526

DO YOU WANT THE 95% CONFIDENCE INTERVALS FOR Y ESTIMATE AND THE PARAMETERS ?YES

95% CONFIDENCE INTERVALS FOR Y AND Y*

Y = B0 + B1*(X1-X1BAR) + B2*(X2-X2BAR) + ...

Y*= B0 + B1*(X1-X1BAR) + B2*(X2-X2BAR) +... + E

(E IS THE ERROR ON Y)

Y ESTIMATE	Y LOWER LIMIT	Y UPPER LIMIT	Y* LOWER LIMIT	Y* UPPER LIMIT
1.08400	.52001	1.64799	.20509	1.96291
4.84400	4.54253	5.14547	4.10556	5.58244
4.84400	4.54253	5.14547	4.10556	5.58244
4.84400	4.54253	5.14547	4.10556	5.58244
8.60400	8.04001	9.16799	7.72509	9.48291

95% CONFIDENCE INTERVALS FOR PARAMETERS

	LOWER LIMIT	UPPER LIMIT
B 0	4.54253	5.14547
B 1	.82084	1.05916

DO YOU WANT TO EXIT PROGRAM *INTMLR* ?
YES
STOP